



Piero Carninci^{a,b}, Chiara Medaglia^b

<http://dx.medra.org/10.17374/CI.2026.108.3.58>

^aRIKEN, Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045 Japan

^bFondazione Human Technopole, Milano
piero.carninci@fht.org

LA RIVOLUZIONE DEGLI RNA NON CODIFICANTI

La scoperta degli RNA non codificanti ha stravolto la visione della regolazione genica. Sebbene non codifichino proteine, questi RNA modulano l'espressione dei geni interagendo con DNA e RNA, contribuendo così all'organizzazione della cromatina. Il progetto FANTOM6 studia sistematicamente le loro funzioni e interazioni, aprendo nuove prospettive per biomarcatori e terapie basate su RNA.



Negli anni Novanta il passaggio dalla genetica classica alla genomica funzionale trasforma radicalmente la biologia molecolare. Prende avvio il Progetto Genoma Umano (HGP, 1990-2003) che si prefigge di sequenziare l'intero genoma, identificarne tutti i geni e crearne una mappa fisica e genetica [1]. Quando, dopo 13 anni di lavoro, il sequenziamento è completato, il risultato è sorprendente: solo l'1-2% del genoma codifica proteine. Il resto? Molti ipotizzano che sia DNA biochimicamente inattivo detto "spazzatura". Ma qualcosa non quadra, è difficile credere che l'evoluzione abbia preservato per milioni di anni una quantità

così imponente di materiale non funzionale. Si teorizza che se il numero di geni non spiega la complessità umana, allora il vero fattore decisivo risiede nel sofisticato controllo dell'espressione genica. Diventa essenziale comprendere quali siano gli elementi funzionali del genoma, cioè tutte quelle sequenze che non codificano proteine, ma stabiliscono dove e quando i geni vengano attivati o silenziati. Da questa necessità nascono i consorzi di genomica funzionale ENCODE (Encyclopedia of DNA Elements) e FANTOM (Functional Annotation of Mammalian Genome). ENCODE, gestito dal National Human Genome Research Institute ne-

gli Stati Uniti, è volto ad identificare tutti gli elementi regolatori del genoma. Il progetto FANTOM, concepito e coordinato dal RIKEN, in Giappone, è invece incentrato sull'analisi funzionale del trascrittoma dei mammiferi. FANTOM è ora alla sesta edizione e nel corso degli anni si è sviluppato progressivamente, passando dalla creazione di un catalogo dei singoli elementi prodotti dal genoma (mRNA, geni, ed RNA non codificanti) allo studio della regolazione della trascrizione ed alla comprensione del sistema trascrizionale, ovvero la rete di regolazione dei sistemi viventi. I dataset e gli atlanti trascrizionali prodotti dal consorzio FANTOM

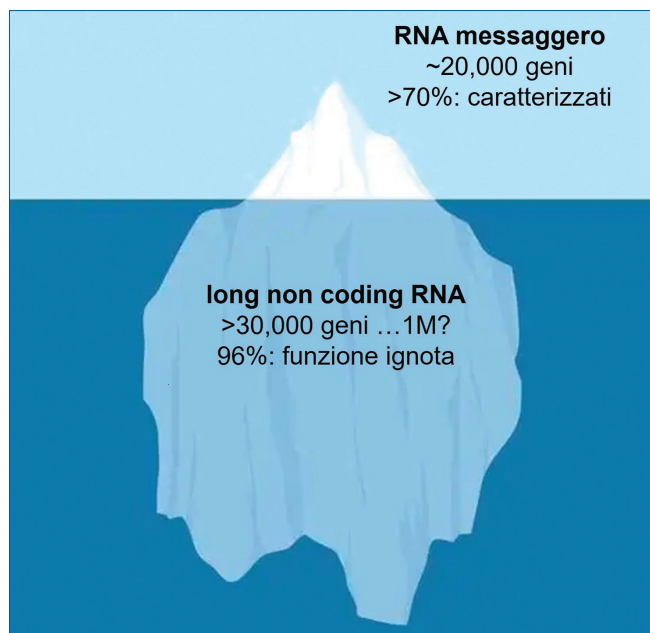


Fig. 1 - Scoperta dei long non-coding RNAs (FANTOM-3, ~2005)

sono assolutamente strumentali all'interpretazione dei dati generati dallo HGP.

L'edizione decisiva del progetto, che ci ha fornito una mappa dinamica dell'attività dei geni, è stata quella del FANTOM 3 (2004-2009), con l'introduzione della tecnologia CAGE (Cap Analysis of Gene Expression) [2]. Il CAGE consente di identificare con precisione i punti di inizio della trascrizione di un gene (TSS, Transcription Start Site) e di misurarne il livello di espressione. Questa metodologia sfrutta la presenza, all'estremità dell'mRNA, del 5' cap, una 7-metilguanosiina aggiunta subito dopo l'inizio della trascrizione e che indica con precisione il TSS. Nel CAGE si estrae l'RNA totale da cellule o tessuti e si catturano solo gli mRNA che possiedono il 5' cap.

Gli RNA vengono poi convertiti in cDNA e sequenziati. La porzione di cDNA contenente il 5' cap è chiamata CAGE tag e la sua posizione esatta sul genoma indica la localizzazione precisa dei promotori, mentre il numero

di tag in quella posizione indica il livello di espressione genica. L'impatto rivoluzionario di CAGE non è dovuto alla mera identificazione dei TSS, ma alla scoperta di quanto fosse pervasiva e sistematica la trascrizione non codificante nel genoma (~80%). Questo dato ha rivelato l'abbondanza dei long non-coding RNA (lncRNA), trascritti che non vengono tradotti. Prima di FANTOM3 i lncRNA noti erano pochi e considerati casi particolari, studiati senza una visione d'insieme. Grazie a CAGE ed altre tecnologie sviluppate dal RIKEN per catturare i cDNA full-length, i lncRNA sono stati identificati come una nuova classe di trascritti e successivi studi hanno dimostrato il loro ruolo nella regolazione biologica. Se ne contano più di 27.000 nell'uomo [3], ma probabilmente sono molti di più (Fig. 1). I lncRNA hanno scardinato il dogma "DNA→RNA→Proteina", secondo cui l'informazione contenuta nel DNA viene utilizzata esclusivamente per produrre proteine. Questo ha cambiato profondamente la visione del genoma, non più inteso come una lista di geni, ma una rete di regolazione. La storia della loro scoperta, avvenuta per caso, ci ricorda l'importanza della serendipità nelle scoperte scientifiche. Nel 1854 Louis Pasteur affermava che *la fortuna aiuta le menti preparate*. È vero, la serendipità non è semplice fortuna, bensì la capacità di riconoscere l'innovazione e l'opportunità dove inizialmente vi era solo un'anomalia. E noi dobbiamo cogliere l'opportunità di questa scoperta.

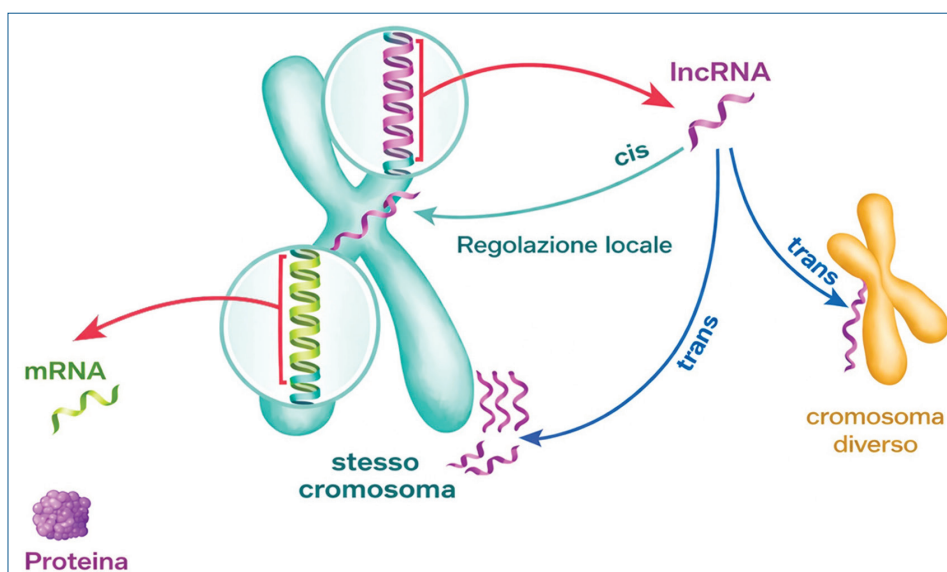


Fig. 2 - Meccanismo d'azione dei long non-coding RNAs (lncRNA)

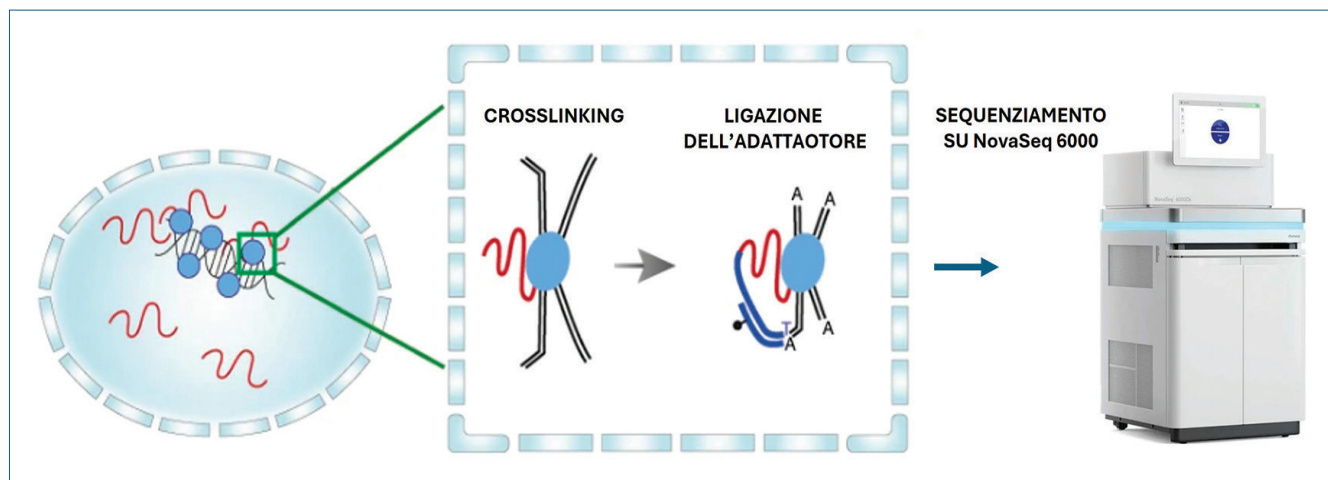


Fig. 3 - Rappresentazione schematica di RADICL-seq. Dopo l'isolamento dei nuclei e il crosslinking, gli ibridi RNA-cromatina (in rosso e in nero, rispettivamente), vengono legati da un adattatore (in blu). Il complesso viene poi sequenziato ad alta risoluzione. L'immagine è stata adattata da A. Bonetti *et al.*, *Nat. Comm.*, 2020, 11, 1018

Cercavamo gli mRNA codificanti; invece, abbiamo trovato una quantità inaspettatamente maggiore di lncRNAs!

C'è ancora molto da capire sui lncRNA (Fig. 2).

Quanti sono? Interagiscono con la cromatina in *cis* o in *trans*? Attivano o reprimono geni? **FANTOM6** (2020-oggi) cerca risposte a queste domande.

Nella sua prima fase, di screening funzionale, sono stati utilizzati oligonucleotidi antisense (GapmeR) per inibire in maniera specifica l'espressione di centinaia di lncRNA in fibroblasti dermici e cellule staminali pluripotenti indotte (iPSC). Gli effetti del silenziamento dei lncRNA sono stati poi misurati tramite imaging in tempo reale e profilazione trascrittomica basata su CAGE. Questo approccio ha portato a capire che una parte significativa dei lncRNA regola processi come crescita, morfologia e migrazione delle cellule [4-7].

La seconda fase, di interattomica, è focalizzata sulla comprensione delle funzioni dei lncRNA attraverso lo studio delle loro interazioni fisiche con il genoma ed il trascrittoma. Utilizzando CAGE e il sequenziamento long-read è stato innanzitutto generato un catalogo completo di ncRNA ed eRNA in un ampio pannello di cellule umane. Sono stati così identificati nuovi elementi regolatori e lncRNA implicati nel differenziamento delle iPSC in neuroni corticali e dei monociti in macrofagi. Poi si è proceduto a costruire una mappa delle interazioni fisiche RNA-DNA, RNA-RNA e DNA-DNA. Sono state ideate *ad hoc*

delle tecnologie progettate per analisi interattomiche su larga scala: **RADICL-seq** (sviluppata dal RIKEN) e **PARIS-seq**, i cui dati vengono poi integrati con le mappe della struttura 3D della cromatina. RADICL-seq (RNA And DNA Interacting Complexes Ligated and sequenced) permette di mappare su scala genomica le interazioni, tra RNA e cromatina. L'idea di base è individuare il punto in cui un RNA si lega o si localizza sul genoma. Le cellule da analizzare vengono sottoposte a crosslinking per fissare i complessi RNA-DNA. Poi i nuclei vengono isolati e la cromatina frammentata. A questo punto, mediante l'utilizzo di un adattatore-ponte, molecole di RNA e DNA che si trovano in prossimità vengono legate irreversibilmente e poi sequenziate (Fig. 3). Abbiamo scoperto che le interazioni RNA-cromatina sono dinamiche e specifiche in base al tipo cellulare. Identificare tali interazioni permette quindi di capire quali RNA regolano i meccanismi alla base della differenziazione cellulare e dell'identità cellulare. Inoltre, RADICL-seq fornisce informazioni sul meccanismo d'azione dei lncRNA, sia in *cis* che in *trans*.

PARIS-seq (Psoralen Analysis of RNA Interactions and Structures) è una tecnologia sviluppata per identificare interazioni RNA-RNA all'interno del nucleo e strutture secondarie e terziarie dell'RNA. Il procedimento tecnico è simile a quello del DICL-seq, con la differenza che porta alla formazione di molecole chimeriche derivanti da interazioni RNA-RNA, che vengono poi sequenziate.



Ora, sappiamo che nel nucleo il DNA è ripiegato ed organizzato in una struttura tridimensionale complessa che influenza le relazioni tra geni ed elementi regolatori. Quindi le mappe di interazione RNA-cromatina ed RNA-RNA vanno inserite nel contesto della disposizione spaziale del DNA che non solo è dinamica, ma può variare tra diversi tipi cellulari ed in risposta a stimoli ambientali. I lncRNA che interagiscono con elementi regolatori spesso modulano la funzione di una certa regione, o “locus”.

Per questo motivo i dati di RADICL-seq e PARIS vengono poi integrati con quelli di interazione DNA-DNA ottenuti mediante **Hi-C**, che permette di capire come le interazioni spaziali tra diverse regioni del DNA, anche se distanti lungo la sequenza, regolano l'attività genica. Hi-C si basa sulla ligazione ed il successivo sequenziamento di frammenti di DNA che si trovano in prossimità nello spazio. I dati ottenuti consentono di identificare compartimenti genomici, ossia regioni genomiche attive o inattive, domini genomici TADs (Topologically Associating Domains) in cui le interazioni sono più frequenti e loop cromatinici, cioè contatti specifici tra elementi regolatori come enhancer e promotori. Si ottiene così una panoramica completa di quali trascritti entrano in contatto con la cromatina ed in quali domini.

Infine, per interpretare queste mappe interattomiche nell'ambito di un contesto fisiologico, i risultati vengono integrati con analisi ottenute da diverse tecnologie di genomica funzionale ad alta risoluzione come **scRNA-seq** (analisi dell'espressione genica a livello di singola cellula), **ATAC-seq** (analisi dell'accessibilità della cromatina), **CUT&Tag** (mappatura di modificazioni istoniche della cromatina) e screening genetici basati su **CRISPRi** (perturbazione di specifici geni o elementi regolatori). Queste metodologie analizzano diversi livelli della regolazione genica, fornendo un quadro integrato dei meccanismi che governano la cellula.

Il progetto si avvia verso la conclusione di questa fase esplorativa, ed i dati ottenuti sono complessi e sorprendenti allo stesso tempo. Alcuni RNA interagiscono con la cromatina all'interno dei TADs, altri invece no. Ci sono RNA che legano il DNA in prossimità del sito di trascrizione, altri invece contattano regioni più lontane del nucleo. Inaspettatamente, alcuni RNA che interagiscono con la cro-

matina derivano da introni. Le interazioni variano in base al tipo di cellule e seguono pattern specifici durante i processi infiammatori o di differenzamento. Se visualizzati con tecniche di imaging, gli RNA che interagiscono con la cromatina appaiono come condensati e confinati in “RNA clouds”, compartimenti sub-cellulari la cui funzione non è ancora del tutto nota. Ovviamente c'è ancora molta strada da percorrere. Resta da comprendere, ad esempio, come misurare queste interazioni all'interno di singole cellule, per poter catturare appieno l'eterogeneità dei tessuti. Questa è una sfida avvincente che abbiamo intrapreso in collaborazione con lo **Human Cell Atlas**. Le scoperte apportate dal FANTOM stanno rivoluzionando il modo in cui studieremo la cromatina. In futuro gli RNA che interagiscono con essa verranno utilizzati per progettare terapie e strategie diagnostiche, e creeranno risorse aperte alla comunità scientifica per disegnare nuovi studi funzionali del genoma.

BIBLIOGRAFIA

- [1] International Human Genome Sequencing Consortium, *Nature* 2004, **431**, 931.
- [2] P. Carninci *et al.*, *Science*, 2005, **309**, 5740.
- [3] C.C. Hon *et al.*, *Nature*, 2017, **543**, 199.
- [4] C.W. Yip *et al.*, *Cell Rep.*, 2022, **41**, 111893.
- [5] J.A. Ramilowski *et al.*, *Genome Res.*, 2020, **30**, 1060.
- [6] S. Agrawal *et al.*, *PLOS ONE*, 2024, **19**, e0295971.
- [7] T. Nobusada *et al.*, *Nucleic Acids Res.*, 2025, **53**, D419.

The Revolution of Long Non-Coding RNAs

The discovery of long non-coding RNAs has profoundly reshaped our understanding of gene regulation. Although they do not encode proteins, many of these RNAs modulate gene expression by interacting with DNA and RNA and contribute to chromatin organization. The FANTOM6 project investigates their functions and interactions, opening new perspectives for RNA-based biomarkers and therapeutic strategies.